

e-Styleistシステムにおける感性用語を用いた商品推薦機能

学習院大学経済学部 白田 由香利

あらまし 本研究では、データマイニングの手法を用いた、eショッピングサイトにおける実現可能かつ有効な商品推薦機能について論じる。我々が研究しているe-Styleistシステムとは、洋服販売サイトにおいて消費者の嗜好に合った洋服を選択し提示してくれるシステムである。このe-Styleistシステム上で、感性用語を用いた商品推薦を行う機能をどのように実現するか、その実現方式を本稿では論じる。また、サッカーユニフォームを注文するサイトを具体的対象とし、感性用語を用いた商品推薦機能を試作したので、それを報告する。本推薦機能の特長は消費者、特に若い消費者が自分の新しい感性用語でフィルタリング可能な点である。サービス提供者側はその商品に特有な良い点を表現する正の評価指標となる、魅力的かつトレンド的な感性用語を定義し、商品推薦の際に利用可能とする。消費者はその感性用語を利用して商品を選択することが可能となる。提案する感性用語の定義方式においては、客観性のある複数の指標から影響を与えている指標を判別し回帰式を求め、これを定義式とする。次に未知なる商品に対して、この回帰式から感性用語に関する指標値を計算する。

キーワード eショッピング、感性用語、推薦システム、データマイニング、回帰分析、要因選択基準

1. 始めに

インターネットの普及に伴い、WEBを介してのeショッピングシステムの数が増加している。こうしたeショッピングシステムは品数が豊富であり消費者にとって利便性が高い。しかしデザインが重要となる品物の購入の際には、購入する品物を選択決定する作業が煩雑となってしまう、という課題がある。こうした問題を解決するためには、高機能なeショッピング用ナビゲーション・システムが必要となる。著者が目標とするナビゲーション・システムとは、婦人服のようなデザインに強く関連する商品を購入する際、支援してくれるシステムである。その仕様は、多数のHTMLあるいはXMLの形式のカatalogを入力すると、消費者の嗜好とセンス、及び購入条件に合った商品を内部評価機能によって選択し、それらのHTML/XML Catalogを再構成し、見易い形式で消費者に呈示してくれる機能をもつものである。

このシステムをe-Styleistと呼ぶことにする。

実世界のスタイリストとは、予算内でイメージに合う洋服を選んでくれたり、小物まで含めてトータルにファッションをチェックしてくれたりするが、そのようなスタイリスト機能をマシン上で実現するための技術的ポイントは次のようにまとめられる。

- (a) 知識の収集と整理：HTMLファイルからの意味構造抽出と表現形式の統一
- (b) 評価及び選択：用途、嗜好等の消費者のニーズに合った商品の評価及び選択
- (c) 商品の提示：HTML/XMLファイル群の再構成
- (d) 対話による商品絞込み：消費者の反応に依存した評価指標の追加・修正を対話的に行う機能

本e-Styleistシステム構築のためには、まず消費

者がどのように商品进行评估し、比較検討を行っているのか、その際の評価プロセスを分析する必要がある。そこでその評価プロセス及び評価属性のモデル化を行ない、次に、その評価属性値をHTMLファイルからどのように抽出したらよいか、また、それを表現できるようなスキーマはどうあるべきか、を考えた^[1]。本e-Stylelistシステムでの最終ゴールは、図1に示す6つの評価を行ってくれるシステムの構築である。これら6種類の評価のうち、洋服選択において最も重要な評価はデザイン評価である。よって、我々の研究はデザイン評価をマシン上で実現する手法から着手した。

本研究では、データマイニングの手法を用いた、e-Stylelistシステムにおける感性用語を利用した商品推薦機能について考察する。本システムモデルの特長は消費者がトレンドな感性用語を利用して商品を選択可能なことである。商品の指標として新たな感性用語を定義するため、客観性のある複数の指標から、影響を与えている指標を判別し回帰式を得る。具体的事例として、サッカーユニフォーム・デザインに対して新しい感性用語「強そう」を指標としてどのように定義するかを示す。さらに結果として得た回帰式から予測を行

ない、その予測の正当性を評価する。また、e-Stylelistシステム中で商品評価が行われた後、これらの商品カタログは消費者にとって、見易く選択し易い形で呈示されるべきである。そこでどのような提示方式が効果的であるかも本稿では論じる。

次節では、eショッピングの洋服選択における課題と考察を、第3節では関連研究について述べる。第4節では、サッカーユニフォーム注文サイトを対象とする推薦システムについて論じる。第5節では、具体的感性用語を新たに定義する手法と、その結果得られた定義式の評価を行なう。第6節はまとめとする。

2. 洋服選択過程に関する考察

本節では、eショッピングにおいて洋服を選択する過程の特徴を考察し、一般的な情報検索及びデータマイニングとの差異を明確化し、洋服の選択に適した推薦機能を探る。

WEBのようにデータ数が膨大な場合、消費者が希望するデータを多少不完全でもフィルタリングし推薦してくれる機能への要望が高い。こうした推薦機能においては、予め対象商品に対して複数の指標（属性）を定義しておき、他方、消費者（達）の過去の購入履歴情報、予め作成したプロフィール情報、あるいは、万人にほぼ共通なデザイン色彩感覚情報などを利用して情報検索あるいはデータマイニングなどの技術により消費者の嗜好に合う商品を絞り込むという手法が採られる。一般的なデータマイニングにおいては、予め用意された膨大なデータベースを対象とし、データ間の相関を発見していくが、上記のようなeショッピングサイトで有用なデータマイニングを考える場合、一般のデータマイニングとの違いに着目する必要がある。

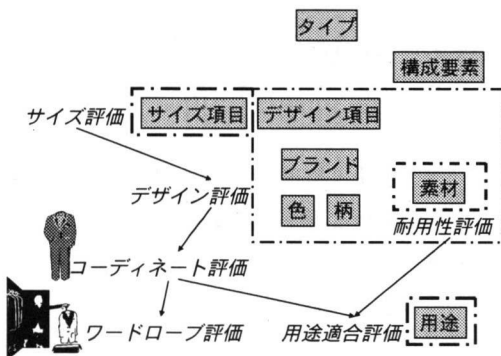


図1 婦人服選択・決定の際の評価プロセスモデル

[1] 白田由香利、インターネット通販におけるカタログ再構成手法—洋服購入における評価指標—、夏のデータベースワークショップ2001、函館、2001-DBS-125-25、pp.185-192、2001年7月17日～19日。

一般的な情報検索及びデータマイニングにおいては、予め用意された膨大なデータベースを対象とし、検索条件を満たす商品や、データ間の相関を発見していく^{[2][3]}。本研究で対象とする洋服のeショッピングサイトのデータベースを考える場合、一般のデータマイニングと比較して以下のような差異がある。

- (a) 新商品の追加を大量かつ迅速に行なう必要があるため、データ入力コストの面から、新商品の評価情報をメタデータとしてデータベースに付加することが十分に行なえない。
- (b) 顧客の入替りがダイナミックであり、かつ、顧客は煩雑なプロフィール作成作業を嫌うため、顧客の情報採取が困難である。
- (c) デザインに関する嗜好は個人に依存する度合いが高いので、一般的なルールは発見が困難。
- (d) 食品などの他の商品に比較して、洋服の購入目的は購入毎の目的が変動する可能性が高い。購入の際の目的に依存した選択指標選択をしないと分析結果に信頼性を欠く。
- (e) デザインは流行があり、特に洋服に関する顧客の選択基準は大幅に変動するため、過去の購買履歴情報などをそのまま長期間利用したのでは、分析結果に信頼性を欠く。

以下では上記の特徴をもつ洋服のeショッピングにおけるデータの特徴に対してどのような方針を採るかを考察していく。

まず(a)であるが、eショッピングサイトでは、検索のためのメタデータの重要度よりも、新商品のデータ登録の、即時性及び入力の低コスト化の方が優先される。よってデータ登録の負荷は小さ

くすべきである。そのためには、スキーマ設計においては、データ入力を必須とする属性は客観性の高いものに限定することが必要である。

上記(b)に関しては、eショッピングサイトの実用性を重視し、顧客の購入履歴、プロフィール情報などは無くても稼動するようにすべきであろう。そのため、感性用語を用いた推薦機能の実現においても、共通感覚をベースとした方がよい。客観性の高い指標をベースにすることにより、システムのフィルタリング結果の精度向上を図る。これは上記(c)と矛盾するが、感性用語の利用も2階層として、1層目で共通感覚、2層目で個人に特化した感覚によるフィルタリングをすべきである、という主張である。

また上記(d)で述べたように、婦人服購入の目的やコンテキストは毎回異なる。この婦人服選択における特徴は以下のようにまとめられる。

- (1) 仕事とプライベートで選択基準が異なる。
- (2) 冠婚葬祭用ではその人本来の嗜好とは異なる選択基準が存在する。黒服を購入した事例があるからといって、常時黒が好きな色であると推論はできない。プライベートでは赤が好きでも冠婚葬祭用に黒服を購入した可能性もある。
- (3) 選択理由は、商品カタログ情報以外の情報と深く関連する。よって、選択基準はコンテキストベースに意味的に深く分析する必要がある。ブランドの宣伝の印象や、有名人が着ていたからなど、洋服自体のもつ特徴以外の指標が深く関与する^[4]。
- (4) 婦人服選択において影響を与える指標は感性的かつ非論理的である。服選択の自由度

[2] A. M. Hughes, Strategic Database Marketing (Second Edition), McGraw-Hill, New York, 2000.

[3] 福田剛志、森本康彦、徳山豪、データサイエンス・シリーズ3：データマイニング、共立出版、東京、2001.

[4] J. Paul Peter, and Jerry C. Olson, Consumer Behavior and Marketing Strategy (Fifth Edition), Irwin/McGraw-Hill, Boston, 1999.

が高く、消費者本人でも明確なイメージがもてない場合、感性的指標が有効となる。

実際、婦人服購入の際の潜在的指標の数は多く、その洋服決定に影響を与えた指標の特定は容易ではない。よってその婦人服購入の際の選択理由の分析においては、コンテキストの分析が重要であり、顧客の購入履歴をデータマイニングする場合においても、潜在的指標の存在に気づかず、全データを同一目的の元であると仮定して分析してしまうことは、結果の信頼性の低下につながる。しかしこうしたコンテキスト分析を完全に行うためには、必要とするデータマイニングは複雑となり、顧客の購買履歴情報も詳細に採取する必要がある。この課題については次節でも考察するが、商品購入がどのような目的であるかを発見するためのコンテキスト分析を行うアプローチとしては、Adomavicius達の提案している多次元的アプローチなどが有効と考えられる。

次に、我々が想定する感性用語とはどのようなものであるかについて説明する。まずe-Stylelistシステムで考える感性用語とは、消費者にとって魅力的かつトレンドで、消費者の購買意欲を向上させるものである必要がある。

婦人服の場合を考えることにする。婦人服のそのシーズンのトレンドは女性ファッション雑誌などで作られることが多い。これらの雑誌に載っている、ファッション用語は、我々の想定する感性用語のひとつである。例えば、「ピュアでナチュラル」、「優しい曖昧カラー（ページュのようでグレー、グレーのようでピンク）」「シャープ&フェミニンな着こなし」「手作り感のある」などである。

また、時代を表現するような流行語も感性用語として利用する。例えば、「いけてるデザイン」というような感性用語である。また、消費者が真似たいと考える有名人の全体的な雰囲気表現する感性用語として、その人名を使って、「x x x

風」というようなものも広義の感性用語と考える。例えば、「川原亜矢子風」というようなものである。

上記(e)に対しては、洋服デザインの普遍性といった潜在的指標を別途定義する必要がある。普遍性の高い分類の服であればデータの寿命も長い。が、普遍性の無い、流行に左右される服であれば、データの寿命も短いので、データの入れ替えが必要となる。

上記の考察の結果、我々は、将来的には詳細な購買履歴分析を行なう予定であるが、始めは履歴情報無しでも有効に分析が行なえる買い物状況を設定するべきであると考え、サッカーユニフォームを対象商品として選択した。サッカーユニフォームのデザイン選択過程は、購入目的が「サッカーをするために着る」と一義的であり、購入のコンテキストもシンプルである。男性消費者が着るので、女性特有の買い物過程の複雑さも無い。さらにデザインの自由度も一般の洋服に比較して高くなく、かつデザインは流行に左右されにくいという特徴をもつ。

本研究では、サッカーユニフォームのデザイン選択過程を対象とし、サッカーユニフォームのデザイン選択のため、感性用語を利用したシステムモデルを提案する。本システムモデルの特長は、消費者、特に若い消費者が自分の新しい感性用語でフィルタリング可能な点である。例えば、サッカーユニフォームに対して「いけてる」、「強そう」、「歴史がありそう」のような、その商品に特有な「良さ」を表現する正の評価指標となる感性用語を定義が可能なことである。しかし、消費者個人々々による感性モデル作成作業は不要とした。そのため個々の消費者に感性モデル作成を行なわせることなく、商品情報提供者側で共通感覚の感性モデル作成を行なう。

3. 関連研究

前節の最後で述べたような、感性用語を利用し

た推薦機能に関連する研究について述べる。インターネット上でのeコマースに関するデータマイニング自体がまだ始まってから時間がたっておらず^[5]、感性用語を用いた研究は少ない。eコマースに対する一般的データマイニングの利点及び今後の課題についてはKohaviが論じている^[6]。

感性モデルを用いたシステムである矢野達によるレコメンデーションシステムにおいては、消費者ごとのイメージ語ネットを予め作成するために、消費者に商品を見せて、それに対してシステムが質問をする形で主観的イメージ、及び客観的イメージを消費者に回答してもらう形を採る^[7]。その結果例えば「スマートなのは、クールかつソフトだから」というルールを得る。この例では、クール、ソフトという2つの客観的イメージと、スマートという主観的イメージの対応関係ルールを得ている。このシステムでは、新商品が登録されると、店側がその商品に付加した客観的イメージ語（例：「クール、ソフト、丸い」）をエージェントに渡す。エージェントは、消費者の個々の感性モデルから、「クール、ソフト」を共通してもつネットワークを探し、あれば、この商品に対してこの消費者は興味をもつと推測する。

矢野達のシステム^[7]では、消費者が自分で主観的イメージを定義する必要があるが、実際のeショッピングサイトにおいて、主観的イメージをこ

のような対話形式で定義する作業を消費者が行なってくれる否か、筆者は大いに疑問がある。少なくともルールを定義していると感じさせないインタフェースが必要と考える。また矢野達のシステムにおいては、客観的イメージを用いて、さらに主観性の高いイメージを定義するという手法を採っているが、ベースとなる客観的イメージを選択する段階において、既に主観的評価が含まれてしまうと考える。

また、クール、ソフトの2つの指標は、デザインのイメージを表現する代表的かつ重要な指標であるが、この2個の指標によるイメージ語の分類研究としては小林による研究蓄積があり、カラーマーケティング戦略の分野で広く利用されている^{[8][9][10][11]}。この研究成果の方が一般性が高く、ルール定義に不慣れな利用者による定義よりも信頼性も高いと考える。

矢野達のアプローチと比較して、我々のアプローチには以下のような特長がある。

- ・ 感性用語の定義に使うベース指標として感性用語は利用せず、客観的な物理的デザイン特徴のみとする。それにより結果の信頼性向上を図る。
- ・ 消費者によるプロフィール作成作業は不要とする。
- ・ 消費者が買い物過程を楽しめるような、流行

[5] Ron Kohavi and Foster J. Provost, Applications of Data Mining to Electronic Commerce, Data Mining and Knowledge Discovery, Kluwer Academic, pp. 5-10, Vol. 5, No. 1/2 (Special Issue), Jan. 2001.

[6] Ron Kohavi, Mining e-commerce data: the good, the bad, and the ugly, Proc. of KDD 2001(International Conf. on Knowledge Discovery and Data Mining), pp. 8-13, San Francisco, CA., August 26-29, 2001.

[7] 矢野絵美、北野有亮、末吉恵美、篠原勲、ピンヤボン シニーナット、加藤俊一、消費者の感性モデルを利用したレコメンデーションシステムの構築、Proc. of DBWeb2002, 情報処理学会シンポジウムシリーズ Vol.2002, No.19, pp.283-289、東京、2002年12年3-4日。

[8] 小林重順、カラーマーケティング戦略、小林重順（編）、（社）日本能率協会、東京、1988。

[9] 小林重順、カラーリストー色彩心理ハンドブッカー、講談社、東京、1997。

[10] 小林重順、配色イメージワーク、講談社、東京、1995。

[11] 道江義頼、室田理子、インテリアとカラーコーディネート、山海堂、東京、1992。

感覚を取り入れたセンスのよい、かつその商品に適した感性用語をシステム側で定義し、消費者が商品検索に利用可能とする。これにより、他のWEBサイトとの差別化を図る。

次に一般的な推薦システムについて簡単に述べる。推薦システムの実現方法は大きく協調フィルタリングによるものと、コンテキストベースによるものの、2種類ある^{[12][13]}。最初に提案された推薦システムのTapestry^[14]の他、GroupLens^[15]、PHOAKS^[16]などが協調フィルタリングを用いている^{[17][18]}。協調フィルタリングにより、多数の利用者間の相関係数を重みとし、新たな商品に対す

る予測評価などが行なえる。もうひとつはコンテキストベースな手法による推薦機能の実現である^{[19][20][21]}。コンテキストベースなアプローチでは、例えば高い評価点のついた商品の説明文中の、キーワードのTF-IDF^[22]を基にユーザプロフィールを作る。また、これらの2つの手法を組み合わせたアプローチもある^{[23][24]}。

しかしながら、我々が対象とする洋服デザインの推薦システムにおいては、既存推薦システムのような単に商品を選択して推薦する機能だけでは不十分である。商品購入がどのような目的であるかによって、そのコンテキストに適した洋服を選

^[12] Paul Resnick, Hal R. Varian, Recommender Systems - Introduction to the Special Section, CACM, Vol. 40, No. 3, pp. 56-58, March 1997.

^[13] Gediminas Adomavicius and Alexander Tuzhilin, Using Data Mining Methods to Build Customer Profiles, Computer, Vol.34, No.2, pp.74-82, Feb. 2001.

^[14] David Goldberg, David Nichols, Brian M. Oki, and Douglas B. Terry, Using Collaborative Filtering to Weave an Information Tapestry, CACM, Vol. 35, No. 12, pp.61-70, Dec. 1992.

^[15] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl, GroupLens: Applying Collaborative Filtering to Usenet News, CACM, Vol. 40, No. 3, pp. 77-87, March 1997.

^[16] Loren G. Terveen, William C. Hill, Brian Amento, David McDonald, and Josh Creter, Phoaks: A System for Sharing Recommendations, CACM, Vol. 40, No. 3, pp. 59-62, March 1997.

^[17] William C. Hill, Larry Stead, Mark Rosenstein, and George W. Furnas, Recommending and Evaluating Choices in a Virtual Community of Use, Proc. of CHI-95, pp. 194-201, Denver, CO., 1995.

^[18] Upendra Shardanand, Pattie Maes, Social Information Filtering: Algorithms for Automating "Word of Mouth", Proc. of CHI-95, pp. 210-217, Denver, CO., 1995.

^[19] Ken Lang, NewsWeeder: learning to filter netnews, Proceedings of the 12th International Conference on Machine Learning, Morgan Kaufmann publishers Inc., pp. 331-339, San Mateo, CA, USA, 1995.

^[20] Michael J. Pazzani, Jack Muramatsu and Daniel Billsus, Syskill & Webert: Identifying Interesting Web Sites, Proceedings of the Thirteenth National Conference on Artificial Intelligence, pp. 54-61, Portland, 1996.

^[21] Raymond J. Mooney and Lorlene Roy, Content-based book recommending using learning for text categorization, Proceedings 5th ACM Conference on Digital Libraries, ACM Press, New York, pp. 195-204, San Antonio, US, 2000.

^[22] 徳永健伸, 情報検索と言語処理, 東京大学出版会, 東京, 1999.

^[23] Marko Balabanovic, Yoav Shoham, Fab: Content-Based, Collaborative Recommendation, CACM, Vol. 40, No. 3, pp. 66-72, March 1997.

^[24] Michael J. Pazzani, A Framework for Collaborative, Content-Based and Demographic Filtering, Artificial Intelligence Review, Vol. 13, No. 5-6, pp. 393-408, Dec. 1999.

択する必要がある。そのためにどのようなアプローチをすべきは今後の課題とするが、Adomavicius達の提案している多次元的アプローチ^[25]は有効であると考ええる。また、女性のeショッピングに特有な傾向に関する考察としてはSlyke達による研究がある^[26]。婦人服のeショッピングサイトにおける女性購入者のトランザクション調査として有益である。

4. 感性用語を利用した洋服デザイン推薦システム

第2節の洋服選択過程に関する考察から、「婦人服選択において影響を与える指標は感性的かつ非論理的であり、服選択の自由度が高く、消費者本人でも明確なイメージがもてない場合、感性的指標が有効となる」という結論を得た。それを受けて、本節では感性用語を利用した洋服デザイン一般に対する商品推薦システムを提案する。

図2に我々が提案する感性用語を利用した推薦システム概要を示した。まずサービス提供者側は予め商品データベースを構築する。その際、商品の属性は商品の物理的特徴を表す属性のみとし、感性指標に関するデータを入力する必要はない。新しい感性指標を定義する必要がある場合のみ、図2左側に示す新感性指標の定義の作業をサービス提供者側が以下の手順で行なう。まず、商品データベースから代表的なデータをサンプリングする。そのサンプル商品データに対して新感性指標値を手で評価し入力する。次にそのデータを用いてデザイン最適モデルを作成し、その最適モデルを基に重回帰分析を行い、結果として回帰式を得る。この回帰式が感性指標の定義式とな

る。この定義プロセスについては次節で詳細に説明する

既にデータベースに登録した商品データにはまだ新感性指標値は入力されていないので、既存商品に対して求めた回帰式を使って新感性指標値を計算し、登録する。その後、新たな商品データ登録の際、物理的属性のみを手で入力。定義済みの感性指標に関しては回帰式を用いて自動的に計算し、付加する。

次に消費者側の説明を行なう。消費者がWEBで買い物を行なう際、システム側は既に定義されている感性指標の提示を行なう。消費者はそこから希望の感性指標を選択する。システムは指定された感性指標の値の高いものを選択し、買い物候補として消費者に提示する。

5. 感性用語指標の定義手法と評価

本節では、具体的な商品に対して、前節で提案したシステムの処理手続きを詳細に説明するとともに、提案手法が有効であることを検証してい

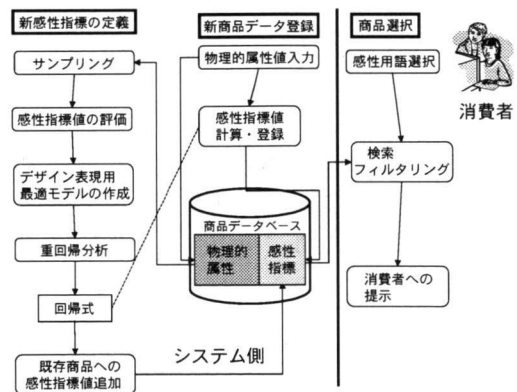


図2 感性用語を利用した洋服デザイン推薦システム

^[25] Gediminas Adomavicius, Alexander Tuzhilin, Multidimensional Recommender Systems: A Data Warehousing Approach, Proc. of Electronic Commerce, Second International Workshop, WELCOM 2001, LNCS 2232, Springer, pp.180-192, Heidelberg, Germany, November 16-17, 2001.

^[26] Craig Van Slyke, Christie L. Comunale, France Belanger, Gender differences in perceptions of web-based shopping, CACM, Vol. 45, No. 8, pp. 82-86, July 2002.

く。本節では、サッカーユニフォーム・デザインに対して、新しい感性用語指標「強そう（に見える）」を定義する手法を示していく。サッカーユニフォームを対象として選択した理由は以下の通りである。まず、第3節で述べたように、婦人服のようなファッション性の高い洋服は選択過程が複雑であり、分析が難しく、複雑なコンテキスト分析なくしては信頼性のある結果は得られない。よってデザイン評価における曖昧性を押さえ、結果の信頼性が保持できるようにするため、一般的なファッション性の高い洋服ではなく、サッカーユニフォームというデザインのシンプルかつ、利用場面がサッカーをするために着るという用途が一意に決まる商品を研究対象として選択した。

本節では、サッカーユニフォームに対して「強そう」という感性的指標を定義した後、結果として得られた用語定義式の予測精度を評価する。

5.1. デザイン表現の最適モデル

サッカーユニフォームのデザインを表現する物理的指標として、以下の4つを選択した。これらはデータベースに属性として格納してある。

- ・ 色1： 第1番目の色。
- ・ 色2： 第2番目の色。
- ・ 柄： ユニフォームの柄
- ・ 色数プラスアルファ： 使っている色数及び、袖襟に他の色が使っているか否か。

上記の属性の取り得る属性値を表1に示す。

サッカーユニフォームのサンプルとして集めた39データに対して、上記の属性値を入力する。こ

表1：サッカーユニフォームデザインモデルのため選択した要因(属性)

色1	色2	柄	色数 + α
白	白	無	単色
青	青	横縞太	2色
赤	赤	縦縞太	袖襟
黄	黄	横縞細	
緑	緑	縦縞細	
黒	黒		

れが新しい指標「強そう」を定義するための物理的特徴値である。次に、上記39データに対して「強そう」の点数づける。評価はサービス提供者側の作業者が行なう。本評価では、評価者は一人である。評価者の選択については本節最後で考察する。

表1に示す複数の属性値を説明変数として、重回帰分析を行なう。重回帰式においては、指標「強そう」に本当に影響を与えている説明変数のみを採用する必要がある。最適な重回帰モデルを求めることは重回帰分析において重要なポイントである。重相関係数Rは重回帰式の良さを表す指標であり、一般にR値が大きいほどよい回帰モデルと言えるが、説明変数を増やし過ぎては最適なモデルといえなくなる。

最適な回帰モデル判別のためには、説明変数選択の際の選択基準として、重相関係数Rをデータ数と説明変数の個数で調整した基準が各種ある^{[27][28]}。我々は、今回、上田が提案する以下の要因選択基準式を利用した^{[29][30]}。

^[27] 佐和隆光、計量経済学の基礎、東洋経済新報社、東京、1970。

^[28] 奥野忠一、芳賀敏郎、矢島敬二、奥野千恵子、橋本茂司、古河陽子、多変量解析法、日科技連出版社、東京、1971。

^[29] 上田太郎、相関があるかを見つける簡便法、オペレーションズ・リサーチ、pp.493-446、1997年7月号。

^[30] 上田太郎、データマイニング事例集、共立出版、東京、1998。

$$\text{要因選択基準} = 1 - (1 - R^2) \times \frac{(\text{データ数} + k + 1)}{(\text{データ数} - k - 1)}$$

上式において、 k はカテゴリ総数、 R は回帰分析における重相関係数を表す。すべての説明変数の組合せに対して要因選択基準を計算し、要因選択基準値が最大となる組合せを最適モデルとして採用する。

表2において○印は組合せを示している。全組合せ数は15通りとなるが、そのうち色1が指定されていないで色2のみが指定されている場合は、評価する意味がないので、その組合せは省略してある。

表2に各説明変数の組合せに対する要因選択基準値の値を示した。 R の値だけを見ると、4つの因子のすべてを含む組合せが最大であるが、要因選択基準値を見ると、(色1、色2、柄)の3属性の組合せが最大値0.04992となっていた。よって最適モデルとして、(色1、色2、柄)の3属性の組合せを採用した。

5.2. 回帰式

このモデルに対して回帰式を計算した結果が以下である。

$$\text{指標「強そう」} = 5.32 + \begin{bmatrix} 0(\text{白}) \\ 2.48(\text{青}) \\ 1.93(\text{赤}) \\ 1.06(\text{黄}) \\ 1.34(\text{緑}) \\ 3.41(\text{黒}) \end{bmatrix} + \begin{bmatrix} 0(\text{白}) \\ -0.13(\text{青}) \\ -0.60(\text{赤}) \\ 0.61(\text{黄}) \\ 1.69(\text{緑}) \\ 2.05(\text{黒}) \end{bmatrix} + \begin{bmatrix} 0(\text{無}) \\ -0.42(\text{横縞太}) \\ 1.32(\text{縦縞太}) \\ 2.81(\text{縦縞細}) \end{bmatrix}$$

分析において説明変数が多過ぎて回帰分析が行えなかったため、かつ、柄が横縞細いであるようなユニフォームデザインは、少なくとも集めたサンプル中には無かったため「柄が横縞細い」という値を外しても問題なしと判断して、変数値から外した。レベル強そうの値域は0～10とした。

表2 選択した要因と要因選択基準値

色1	色2	柄	色数+ α	k	R ²	要因選択基準値
○	○	○	○	15	0.556906	-0.0596
○	○	○	-	13	0.551521	0.04922
○	○	-	○	12	0.522814	0.04563
○	○	-	-	10	0.458344	0.03276
○	-	○	○	10	0.361683	-0.1399
○	-	○	-	8	0.342884	-0.0514
○	-	-	○	7	0.342219	0.00272
○	-	-	-	5	0.266932	0.00036

5.3. 予測と評価

上記結果の回帰式を用いて指標「強そう」の予測を行なう。例えば、青と白色の配色で、太い縦縞柄のユニフォームの予測値は、 $5.32 + 2.48 + 0 + 1.32 = 9.12$ となり、かなり強そうに見えると予測される。

次にこの回帰式の評価を行なう。レベル強そうの採取データ値と予測値の絶対誤差を全データについて計算する。絶対値の平均は16.6%であった。感性を用いた印象度の予測は難しいので、この値は満足できる予測精度と言える。

5.4. 考察

まず、サッカーユニフォームにおける感性的指標「強そう」の定義で用いたデータ採取の方法について考察する。第1の課題として、表1に示したデザインモデルは適切であったか否かがある。サッカーユニフォーム柄の別の類別方法としては、襟、袖口、サイドなどに挿入されたラインの入り方による違いなどによる詳細な分類方式が考えられる。デザインの分類方式は、今後デザインの専門家の意見を聞き、効果的な分類方法を選択すべきであろう。第2の課題としては、「強そう」の点数評価者の選択方針があげられる。今回は、評価者は一人とした。デザイン評価は個人的嗜好の偏りが顕著であるため、多人数の平均を単にとっただけでは、意味がある結果は得られない。アプローチとして、(1)ある特定の一人の人の嗜好モ

デルを作成する、(2)ある嗜好傾向をもつグループに対して嗜好モデルを作成する、の2つが考えられる。前者は、有名人スタイリストなどの消費者がよく知っている人の個性に基づきモデルを作成する場合である。また、後者のアプローチとは、例えば、10代から20代前半の女性にターゲットを絞った商品であれば、若い女性の嗜好モデルを形成するように、その年代の女性を評価者グループとして採用する場合である。本稿で論じた評価者1人による評価は、第1のアプローチである。後者のように、評価者の数を増加する場合については、今後の課題としたい。

次に、最適モデルを作成するための、説明変数の選択方式について論じる。本稿では説明変数の選択基準として上田の提案する要因選択基準の手法を採用したが、SPSSで提供する変数選択手法も利用可能である^[31]。SPSSでは、変数選択アルゴリズムとして、変数増加法 (forward)、変数減少法 (backward)、ステップワイズ法、総当り法を提供している。これらの選択方式によって結果が異なることがあるので、利用の際には、後述する俯瞰ツール (図3 参照) などを用いて、外れ値の検討などを行って、結論の正当性をチェックすることが望まれる。

回帰分析を行なう場合、外れ値の発見など、分析結果の正当性を目視でチェックすることは重要である。こうした属性値の入力チェックのため、図3に示すよ

うな、散布図にアイコンをレイアウト表示する俯瞰ツールを利用した。本俯瞰ツールは、商品データベースのデータを用いて、数値属性をもつ任意の2つの属性を指標軸として2次元散布図を描画する機能をもつ。図3では、本節で定義した「強そう (strong)」と、「クール (cool)」の2つの指標を2軸に採ってプロットした。クールの反意語は暖かいであり、図3では、寒色系と暖色系の商品カテゴリの大きな2つの島が形成されていることが分かる。本俯瞰ツールでは、商品データベースに登録する属性としてアイコンの画像ファイル名を格納しておくことにより、散布図の該当位置にアイコン画像を配置できる。また、商品データベースに詳細説明カタログのURLも登録可能であり、アイコン画像を散布図上でクリックすることにより、対応する商品の詳細説明カタログ情報も表示されるようになっている。この散布図ツールにより、デザインを重要要素とする商品の回帰分析の結果の目視チェックが容易になった。

また、消費者がカタログを自分の選択指標により再構成して表示可能である、という点でも本俯

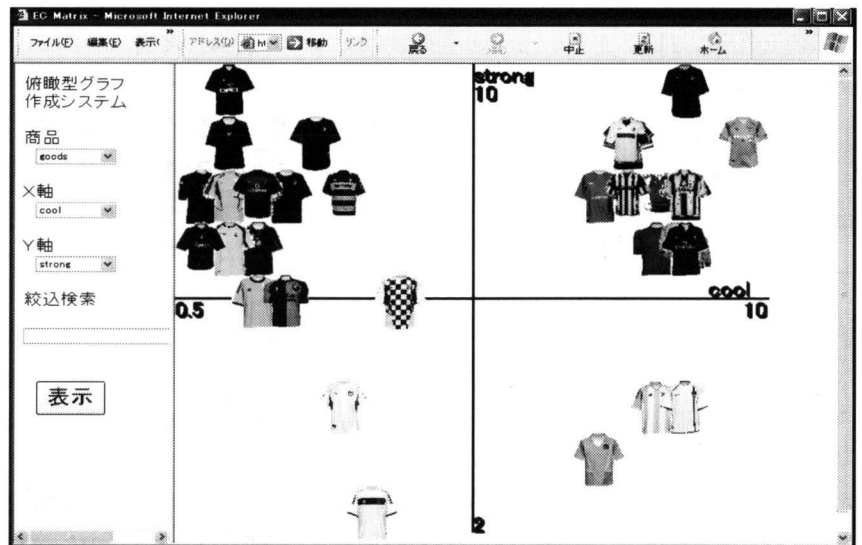


図3 俯瞰ツールによるサッカーユニフォームのデザイン分類

[31] SPSS Inc., http://www.spss.co.jp/product/spss_tables/tables.html

瞰ツールは利便性が高い。デザイン評価以外の一般的な商品選択過程においても有効である。消費者は一般にWebで提示される商品数が膨大な場合、どのような指標があるのか、また、その指標によりどのような商品分布が形成されるのか知りたい、という要望が高い。例えば、デジタルカメラの物理的指標である、画素数と重量を2軸として選択しプロットするような場合を想定する。そのような場合に、このような俯瞰ツールを用いることで、(1)自分が注目する指標（感性的指標を含む）を選択、(2)その指標の値によって、商品群を2次元上に配置という操作が可能となる。

また、洋服のようなデザイン性の高い商品の場合、小さいアイコンでもデザインや色の概観が画面上に表示されることでそれら商品のイメージを把握しやすくなるという利点が生まれる。本俯瞰ツールでは、データベースに格納された各種の属性によって表示する商品を絞り込むことも可能となっている。画面上左側下にある検索条件のテキスト入力エリアに、SQL文の文法に従って検索条件を書くことで、条件に合致した商品だけが表示される。本俯瞰ツールはWeb上で利用可能なので、HTML/XMLカタログとの連携も容易であり、デザインアイコンをクリックすることにより、消費者は興味のある商品の詳細なカタログ情報を現在見ているものと同一のWebブラウザ上に表示することが可能である。

6. まとめ

我々はe-Stylistシステムと言う、洋服販売サイトにおいて消費者の嗜好に合った洋服を選択し提示してくれるシステムについて研究を行っている。洋服、特に婦人服における選択指標とは、感性的かつ非論理的である。そしてWebにおけるeショッピングのように品数が多く、消費者本人でも希望商品に対する明確なイメージがもてない場合、感性的指標が有効となる。そこで、我々は本e-Stylistシステム上で、感性用語を用いた商品推

薦を行う機能をどのように実現するか、その実現方法を論じた。具体的には、サッカーユニフォームを注文するサイトを対象とし、感性用語「強そう」を定義し推薦機能に利用するプロセスを詳細に説明した。

我々が提案した、感性用語に関する新しい指標を定義する本手法は、サッカーユニフォームだけではなく、洋服などのデザイン重視の商品全般に適応可能である。その作業は、(1)デザインの物理的特徴から最適なモデルを選択、(2)最適モデルに対して重回帰分析を行い、回帰式を計算するという2ステップから構成される。本提案手法には以下のような特長がある。

- ・ 感性用語の定義に使うベース指標として感性用語は利用せず、客観的な物理的デザイン特徴のみとした。これにより新たに定義される指標の信頼性向上が図れる。
- ・ 消費者に煩雑なプロファイル作成作業はさせない。サービス提供者側で作成した商品評価情報のみで稼動するようにした。
- ・ 消費者が買い物過程を楽しめるような、流行感覚を取り入れたセンスのよい、かつその商品に適した感性用語をシステム側で予め定義しておき、消費者が商品検索にその感性用語を利用可能とする。これにより、そのWEBサイトの魅力を増加させ他のWEBサイトとの差別化を図る。
- ・ 新たな商品をデータベースに追加する場合、その商品の物理的特徴値のみを入力するだけで済む。感性用語に基づく指標値は、物理的特徴値から予測可能となる。

我々が提案した感性用語を用いた推薦手法は、婦人服のスタイリスト機能として以下のような利用方法も考えられる。

- (a) 個人の手持ちのワードローブに対して、今シーズンのトレンドに合致する洋服があるか否

かを評価する場合。その結果、トレンドにあっている服がどの程度あるかが分かる。

- (b) 上記を商店の在庫商品に対して行くと、在庫のうちトレンドに合致する商品を選択可能。
- (c) 複数のファッションリーダーやファッション雑誌の提言する今シーズンのトレンドの内容が、合致しているか否かを比較検討することが可能。
- (d) 自分の洋服に対する感性が他の人あるいはグループによって定義された感性とどの位差異があるか比較検討することが可能。

現在、白田ゼミでは、日本茶ペットボトルの味に対する感性用語による推薦方式を検討している。その他、e-Stylelistシステムとして、婦人用トップス及びデザイン重視の小物類に対する感性用語の定義の実験を行っている。今後とも広い範囲にわたる商品群に対して、この感性用語による推薦手法が適応可能か検討していきたい。